PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : H04N 13/02, G06T 3/00

A1

(11) International Publication Number: WO 00/13423

(43) International Publication Date: 9 March 2000 (09.03.00)

(21) International Application Number: PCT/US99/19706

(22) International Filing Date: 30 August 1999 (30.08.99)

(30) Priority Data:

60/098,368	28 August 1998 (28.08.98)	US
Not furnished	27 August 1999 (27.08.99)	US

(71) Applicant: SARNOFF CORPORATION [US/US]; 201 Washington Road, CN5300, Princeton, NJ 08543 (US).

(72) Inventors: HANNA, Keith, J.; 212 John Street, Princeton, NJ 08540 (US). BERGEN, James, R.; 57 E. Prospect Street, Hopewell, NJ 08525 (US). KUMAR, Rakesh; 966 Ridge Road, Monmouth Junction, NJ 08852 (US). SAWHNEY, Harpreet; 17 Melville Road, Cranbury, NJ 08512 (US). LUBIN, Jeffrey; 112 W. 72nd Street, New York, NY 10023 (US).

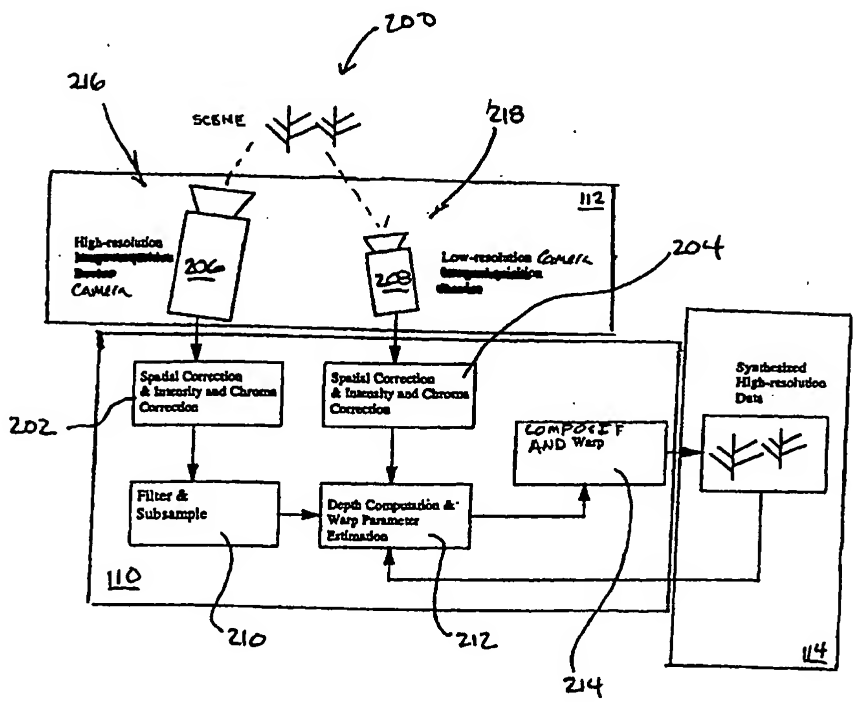
(74) Agent: BURKE, William, J.; Sarnoff Corporation, Law and Patent Operations, 201 Washington Road, CN 5300, Princeton, NJ 08543 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: METHOD AND APPARATUS FOR SYNTHESIZING HIGH-RESOLUTION IMAGERY USING ONE HIGH-RESOLUTION CAMERA AND A LOWER RESOLUTION CAMERA

(57) Abstract



An apparatus and method for transforming imagery recorded by one camera into other imagery that differs from the first imagery, using imagery collected by one or more additional cameras that differ in their characteristics or parameters from the first camera. Example differences include spatial position of the camera, spatial resolution, spectral characteristics and spatial layout. The apparatus generates a synthetic high-resolution image using a high-resolution camera (206) positioned and a lower-resolution camera (208). The high-resolution image data is warped using the lower-resolution data to generate a synthetic high-resolution image (114) having viewpoint of the lower-resolution camera (208). The high-resolution synthetic image generation routine (110) comprises the steps of correcting the spatial, intensity and chromanence distortions of the image data acquired from the high-resolution camera (206) and the lower resolution camera (206) (step 202), subsequently filtering and subsampling the corrected high-resolution data (step 210), computing the parallax between the high-resolution data and the low-resolution data (step 212) and warping the high-resolution image to create a synthetic image (114) of the scene (200) having a viewpoint of the lower-resolution camera (208) (step 214).

METHOD AND APPARATUS FOR SYNTHESIZING HIGH-RESOLUTION IMAGERY USING ONE HIGH-RESOLUTION CAMERA AND A LOWER RESOLUTION CAMERA

CLAIM OF PRIORITY

This application claims the benefit under 35 United States Code § 119 of United States Provisional Application No. 60/098,368, filed August 28, 1998, which is hereby incorporated by reference in its entirety.

This application discloses subject matter that is related to the disclosure in United States patent application number __________, filed simultaneously herewith (Attorney docket no. SAR 13422), which is incorporated herein by reference in its entirety.

The invention relates to an image processing apparatus and, more particularly, the invention relates to a method and apparatus for creating a high-resolution synthetic image from two or more cameras that differ by one or more characteristics or parameters.

BACKGROUND OF THE DISCLOSURE

For entertainment and other applications, it is useful to obtain high-resolution stereo imagery of a scene so that viewers can visualize the scene in three dimensions. To obtain such high-resolution imagery, the common practice of the prior art is to use two or more high-resolution devices or cameras, displaced from each other. The first high-resolution camera captures an image or image sequence, that can be merged with other high-resolution images taken from a viewpoint different than the first high-resolution camera, creating a stereo image of the scene.

However, creating stereo imagery using multiple high-resolution cameras can be difficult and very expensive. The number of high-resolution cameras used to record a scene can contribute significantly to the cost of producing the stereo imagery. Additionally, high-resolution cameras are large and unwieldy. Thus, the ease of which a scene is filmed can be burdensome. Further, some viewpoints may not be able to

accommodate the size of such high-resolution cameras, thus limiting the viewpoints available for creating the stereo image.

Therefore, a need exists in the art for a method and apparatus for creating one or more synthetic images from a plurality of cameras that differ in their characteristics or parameters.

SUMMARY OF THE INVENTION

The disadvantages associated with the prior art are overcome by the present invention of a method and apparatus for transforming imagery recorded by one camera into other imagery that differs from the first imagery, using imagery collected by one or more additional cameras that differ in their characteristics or parameters from the first camera. Example differences include spatial position of the camera, spatial resolution, spectral characteristics and spatial layout. One specific embodiment of the invention is apparatus that comprises a high-resolution camera for producing images at a high-resolution and a lower-resolution camera for producing images at a lower-resolution coupled to an image processor. The image processor Performs various image flow and parallax estimation computations and warps the high-resolution image to a viewpoint of the lower-resolution camera.

The invention includes a method that is embodied as a software routine, or a combination of software and hardware. The inventive method comprises the steps of supplying image data having a high-resolution, supplying image data having a lower-resolution, processing the imagery, then warping the high-resolution image to a viewpoint of the lower-resolution image data to form a synthetic image. As such, the original high-resolution image and the synthetic image can be used to form a high-resolution stereo image using only a single high-resolution camera.

BRIEF DESCRIPTION OF THE DRAWINGS

The teachings of the present invention can be readily understood by considering the following detailed description in conjunction with the accompanying drawings, in which:

Fig. 1 depicts a block diagram of an imaging apparatus incorporating the image analysis method and apparatus of the invention;

Fig. 2 depicts a block schematic of an imaging apparatus and an image analysis method used to produce one embodiment of the subject invention;

Fig. 3 is a flow chart of the parallax computation method; and,

Fig. 4 is a flow chart of the image compositing method.

To facilitate understanding, identical reference numerals have been used, where possible, to designate identical elements that are common to the figures.

DETAILED DESCRIPTION

FIG. 1 depicts a high-resolution synthetic image generation apparatus 100 of the present invention. An input video sequence 112 is supplied to a computer 102. The input sequence 112 may comprise of a pair of frames taken at a single instance, a series of frame pairs taken over time or a series of frames. The computer 102 comprises a central processing unit (CPU) 104, support circuits 106, and memory 108. Residing within the memory 108 is a high-resolution synthetic image generation routine 110. The high-resolution synthetic image generation routine 110 may alternately be readable from another source such as a floppy disk, CD, remote memory source or via a network. The computer additionally is coupled to input/output accessories 118. As a brief description of operation, an input video sequence 112 is supplied to the computer 102, which after operation of the high-resolution synthetic image generation routine 110, outputs a synthetic high-resolution image 114.

An example embodiment of a transform related to the spatial positions of the sensors are the parallax recovery methods described below. An example embodiment of a transform related to the spatial resolution of the sensor is described in Burt and Adelson "Laplacian Pyramid as a compact Image code", where images are transformed from one resolution to other resolutions in the process of computing an image pyramid. An example embodiment of a transform relating to spectral characteristics of the sensors is a mapping from HSL (Hue,saturation,lightness) to RGB (Reg,Green,Blue) as described in "Graphics Gems", edited by Andrew Glassner, Academic Press, 1990. An example of a transform that relates the spatial layout of imagery recorded from one sensor to another spatial layout is described in "A Theory of Catadioptric Image Formation" by S. Baker and S.K. Nayar in the Proceedings of the 6th International Conference on Computer Vision, Pages 35-42, Bombay, India, January, 1998. An additional example of a transform that relates the spatial layout of imagery recorded from one sensor to another spatial layout is described in "An Anthropomorphic Retina-Like Structure for Scene Analysis", by Sandini and Tagliasco, in Journal of Computer Vision, Graphics and Image Processing, Vol 14, p365-372, 1980.

More specifically, the high-resolution synthetic image generation routine 110 hereinafter referred to as the routine 110, can be understood in greater detail by referring to Fig. 2. Although the process of the present invention is discussed as being implemented as a software routine 110, some of the method steps that are disclosed therein may be performed in hardware as well as by the software controller. As such, the invention may be implemented in software as executed upon a computer system, in hardware as an application specific integrated circuit or other type of hardware implementation, or a combination of software and hardware. Thus, the reader should note that each step of the routine 110 should also be construed as having an equivalent application specific hardware device (module), or hardware device used in combination with software.

The high-resolution synthetic image generation routine 110, receives the input 112 from a high resolution camera 206 and a lower

resolution camera 206. The high resolution camera 206 views a scene 200 from a first viewpoint 216 while the lower resolution camera 206 views the scene 200 from a second viewpoint 218. The high resolution camera 206 has an image resolution higher than that of the lower resolution camera 206. The high resolution camera 206 may comprise a number of different devices having a number of different data output formats, as one skilled in the art will readily be able to adapt the process described by the teachings herein to any number of devices and data formats and/or protocols. In one embodiment, the high resolution camera 206 is a high-definition camera, i.e., a model MSM9801 camera, available from IMAX® Corporation. Similarly, the lower resolution camera 206 may also comprise a varied number of devices, since one skilled in the art can readily adapt the routine 110 to various devices as discussed above. In one embodiment, the low-resolution device is a camera having a resolution lower than the resolution of the high-resolution device, i.e., a standard definition video camera. For example, the resolution imagery may have at least 8000 by 6000 pixels/cm^2 and the lower resolution image may have 1000 by 1000 pixels/cm^2.

The routine 110 receives input data from the high resolution camera 206 and corrects the spatial, intensity and chromanence (chroma) distortions in step 202. The chroma distortions are caused by, for example, lens distortion. This correction is desired in order to improve the accuracy of subsequent steps executed in the routine 110. Methods are known in the art for computing a parametric function that describes the lens distortion function. For example, the parameters are recovered in step 202 using a calibration procedure as described in H. S. Sawhney and R. Kumar, *True Multi-Image Alignment and its Application to Mosaicing and Lens Distortion,* Computer Vision and Pattern Recognition Conference proceedings, pages 450-456, 1997, incorporated by reference in its entirety herein.

Chroma and intensity corrections are also performed in step 202. These correction are necessary since image data from the lower resolution camera 206 is merged with data from the high resolution camera 206, and any differences in the device response to scene color and intensity or due to

lens vignetting, for example, results in image artifacts in the synthesized image 114. The correction is performed by pre-calibrating the devices (i.e., the high resolution camera 206 and the lower resolution camera 206) such that the mapping of chroma and intensity from one device to the next is known. The measured chroma and intensity correction information from each device is stored in look-up tables or as a parametric function.

Input data from the lower resolution camera 206 is also corrected for spatial, intensity and chroma distortions in step 204. The process for correcting the low-resolution distortions in step 204 follow the same process as the corrections performed in step 202.

The corrected high-resolution data from step 202 is subsequently filtered and subsampled in step 210 . The purpose of step 210 is to reduce the resolution of the high-resolution imagery such that it matches the resolution of the low-resolution image. Step 210 is necessary since features that appear in the high-resolution imagery may not be present in the lower-resolution imagery, and cause errors in a depth recovery process (step 306 detailed in Fig. 3 below). Specifically, these errors are caused since the depth recovery process 306 attempts to determine the correspondence between the high-resolution imagery and the low-resolution imagery, and if features are present in one image and not the other, then the correspondence process is inherently error-prone.

The step 210 is performed by first calculating the difference in spatial resolution between the high resolution camera 206 and low resolution camera 208. This is performed as a pre-calibration step in which the relative scale of pixels/cm^2 between the two cameras is computed. For example, this relative scale is given by the ratio of lengths or square root of the ratio of areas of a fixed shape that is viewed by the two cameras. From the difference in spatial resolution, a convolution kernel can be computed that reduces the high-frequency components in the high-resolution imagery such that the remaining frequency components match those components in the low-resolution imager. This can be performed using standard, sampling theory (e.g., see P. J. Burt and E. H. Adelson, *The Laplacian Pyramid as a Compact Image Code,* IEEE Transactions on

Communication, Vol. 31, pages 532-540, 1983, incorporated by reference herein in its entirety).

For example, if the high-resolution and low-resolution imagery were different in spatial resolution by a factor of 2 vertically and horizontally, then an appropriate filter kernel is [1,4,6,4,1]/16. The filter kernel is applied first vertically, then horizontally. The high-resolution image can then be sub-sampled by a factor of 2 so that the spatial sampling of the image data derived from the high-resolution imager matches that of the low-resolution imager.

Once the high-resolution image data has been filtered and subsampled in step 210, the parallax is computed in step 212 at each frame time to determine the relationship between viewpoint 216 and viewpoint 218 in the high-resolution and low-resolution data sets. More specifically, the parallax computation of step 212 computes the displacement of image pixels between the images taken from view point 216 and viewpoint 218 due to their difference in viewpoint of the scene 200.

Because this parallax information depends on the relationship between the two input images recorded at a common instance in time and having different viewpoints (216 and 218, respectively) of a scene 200, it is initially computed at the spatial resolution of the lower resolution image. This is accomplished by resampling the high-resolution input image using an appropriate filtering and sub-sampling process, as described above in step 210.

The computation of step 212 is performed using more or less constrained algorithms depending on the assumptions made about the availability and accuracy of calibration information. In the extremely unconstrained case, a two-dimensional flow vector is computed for each pixel in the image for which alignment is being performed. If it is known that the epipolar geometry is stable and accurately known, then the computation reduces to a single value for each image point.

In many situations, particularly those in which parallax magnitudes are large, it is advantageous in step 212 to compute parallax with respect to some local parametric surface. Note that parallax computations are, in effect, a constrained computation of image flow. One

method of parallax computation is known as "plane plus parallax". The plane plus parallax representation can be used to reduce the size of per-pixel quantities that need to be estimated. For example, in the case where scene 200 comprises an urban scene with a lot of approximately planar facets, parallax may be computed in step 212 as a combination of planar layers with additional out-of-plane component of structure. The procedure for performing the plane plus parallax method is detailed in United State Patent Application No. 08/493,632, filed June 22, 1995, R. Kumar et al., *Direct Recovery of Shape From Multiple Views: A Parallax Based Approach*, 12th ICPR, 1994, Harpreet Sawhney, *3D Geometry From Planar Parallax*, CVPR 94, June 1994, and A. Shashua and N. Navab, Relative Affine Structure, *Theory and Application to 3D Construction From 2D Views*, IEEE Conference on Computer Vision and Pattern Recognition, June 1994, all of which are hereby incorporated by reference.

Other algorithms are available that can perform parallax analysis in-lieu of the plane plus parallax method. These algorithms generally use a coarse-fine recursive estimation process using multiresolution image pyramid representations. These algorithms begin estimation of image displacements at reduced resolution and then refine these estimates through repeated warping and residual displacement estimation at successively finer resolution levels. The key advantage of these methods is that they provide very efficient computation even when large displacements are present but also provide sub-pixel accuracy in displacement estimates. A number of published papers describe the underlying techniques employed in the parallax computation of step 212. Details of such techniques can be found in US patent 5,259,040, issued November 2, 1993; J. R. Bergen et al., *Hierarchical Model-Based Motion Estimation*, 2nd European Conference on Computer Vision, pages 237-252, 1992; K. J. Hanna, *Direct Multi-Resolution Estimation of Ego-Motion and Structure From Motion*, IEEE Workshop on Visual Motion, pages 156-162, 1991; K. J. Hanna and Neil E. Okamoto, *Combining Stereo and Motion Analysis for Direct Estimation of Scene Structure*, International Conference on Computer Vision, pages 357-356, 1993; R. Kumar et al., *Direct Recovery of Shape from Multiple Views: A Parallax Based*

Approach, ICPR, pages 685-688, 1994; and S. Ayer and H. S. Sawhney, *Layered Representation of Motion Video Using Robust Maximum-Likelihood Estimation of Mixture Models and MDL Encoding*, International Conference on Computer Vision, pages 777-784, 1995, all of which are hereby incorporated by reference.

Although the step 212 can be satisfied by simply computing parallax using the plane plus parallax method described above, there are a number of techniques that can be used to make the basic two-frame stereo parallax computation of step 212 more robust and reliable. These techniques may be performed singularly or in combination to improve the accuracy of step 212. The techniques are depicted in the block diagram of Fig. 3 and comprise of augmentation routines 302, sharpening routines 304, routines that compute residual parallax 306, occlusion detection routines 308, and motion analysis routines 310. Although these processes are discussed as being useful in improving a parallax computation, the same augmentation processes can be applied to an image flow computation to enhance the accuracy of an image flow estimation.

The augmentation routines 302 make the basic two-frame stereo parallax computation robust and reliable. One approach divides the images into tiles and, within each tile, the parameterization is of a dominant plane and parallax. In particular, the dominant plane could be a frontal plane. The planar parameterization for each tile is constrained through a global rotation and translation (which is either known through pre-calibration of the stereo set up or can be solved for using a direct method). In addition, a single epipolar constraint is applied to all the parallax vectors for any planar tile.

Another augmentation routine 302 handles occlusions and textureless areas that may induce errors into the parallax computation. To process occlusions and textureless areas, depth matching across two frames is performed using varying window sizes, and from coarse to fine spatial frequencies. Multiple window sizes are used at any given resolution level to test for consistency of depth estimate and the quality of the correlation. Depth estimate is considered reliable only if at least two window sizes produce acceptable correlation levels with consistent depth

estimates. Otherwise, the depth at that level is not updated. If the window under consideration does not have sufficient texture, the depth estimate is ignored and a consistent depth estimate from a larger window size is preferred if available. Areas in which the depth remains undefined are labeled as such as to that they can be filled in either using preprocessing, i.e., data from the previous synthetic frame or through temporal predictions using the low-resolution data, i.e., up-sampling low-resolution data to fill in the labeled area in the synthetic image 114. The process for using multiple windows to improve the parallax computation is further disclosed in US patent application serial number ___________, filed simultaneously herewith (Attorney docket no. SAR 13422), which is hereby incorporated by reference in its entirety.

An additional approach for employing an augmentation routine 302 is to use Just Noticeable Difference (JND) models in the optimization for depth estimation. For example, typically image measures such as intensity difference are used to quantify the error in the depth representation. However, these measures can be supplemented with JND measures that attempt to measure errors that are most visible to a human observer. The approach for employing JND methods are discussed in greater detail below.

An additional augmentation routine 302 provides an algorithm for computing image location correspondences. First, all potential correspondences at image locations are defined by a given camera rotation and translation at the furthest possible range, and then correspondences are continuously checked at point locations corresponding to successively closer ranges. Consistency between correspondences recovered between adjacent ranges gives a measure of the accuracy of the correspondence.

Another augmentation routine 302 avoids blank areas around the perimeter of the synthesized image. Since the high-resolution imagery is being warped such that it appears at a different location, the image borders of the synthesized image may not have a correspondence in the original synthesized image. Such areas may potentially be left blank. This problem is solved using three approaches. The first approach is to display only a central window of the original and high-resolution imagery,

such that the problem area is not displayed. The second approach is to use data from previous synthesized frames to fill in the region at the boundary. The third approach is to filter and up-sample the data from the low-resolution device, and insert that data at the image boundary.

An additional augmentation routine 302 provides an algorithm that imposes global 3D and local (multi-) plane constraints Specifically, the approach is to represent flow between frame pairs as tiled parametric (with soft constraints across tiles) and smooth residual flow. In addition, even the tiles can be represented in terms of a small number of parametric layers per tile. In the case when there is a global 3D constraint across the two frames (stereo), then the tiles are represented as planar layers where within a patch more than one plane may exist.

Another method for improving the quality of the parallax computation of step 212 is to employ a sharpening routine 304. For example, in the neighborhood of range discontinuities or other rapid transitions, there is typically a region of intermediate estimated parallax due to the finite spatial support used in the computation process 212. Explicit detection of such transitions and subsequent "sharpening" of the parallax field minimize these errors. As an extension to this basic process, information from earlier (and potentially later) portions of the image sequence is used to improve synthesis of the high-resolution image 114. For example, image detail in occluded areas may be visible from the high-resolution device in preceding or subsequent frames. Use of this information requires computation of motion information from frame to frame as well as computation of parallax. However, this additional computation is performed as needed to correct errors rather than on a continual basis during the processing of the entire sequence.

Additionally, the parallax computation of step 212 can be improved by computing the residual parallax (depth) using a method described as follows or an equivalent method that computes residual parallax 306. One method computes depth consistency over time to further constrain depth/disparity computation when a motion stereo sequence is available as is the case, for example, with a 15-65 formatted hi-resolution still image. Within the two images captured at the same time instant, rigidity

constraint is valid and is exploited in the two-frame computation of depth outlined above. For multiple stereo frames, optical flow is computed between the corresponding frames over time. The optical flow serves as a predictor of depth in the new frames. Within the new frames, depth computation is accomplished between the pair while being constrained with soft constraints coming from the predicted depth estimate. This can be performed forward and backwards in time. Therefore, any areas for which estimates are available at one time instant but not at another can be filled in for both the time instants.

Another method of computing residual parallax 306 is to use the optical flow constraint along with a rigidity constraint for simultaneous depth/disparity computation over multiple stereo pairs. In particular, if large parts of the scene 200 are rigid, then the temporal rigidity constraint is parameterized in the depth computation in exactly the same manner as the rigidity constraint between the two frames at the same time instant. When there may be independently moving components in the scene 200, the optical flow constraint over time may be employed as a soft constraint as a part of the multi-time instant depth computation.

Another method of computing residual parallax 306 is to constrain depth as consistent over time to improve alignment and maintain consistency across the temporal sequence. For example, once depth is recovered at one time instant, the depth at the next frame time can be predicted by shifting the depth by the camera rotation and translation recovered between the old and new frames. This approach can also be extended by propagating the location of identified contours or occlusion boundaries in time to improve parallax or flow computation.

An additional approach for computing residual parallax 306 is to directly solve for temporally smooth stereo, rather than solve for instantaneous depth, and impose subsequent constraints to smooth the result. This can be implemented using a combined epipolar and flow constraint. For example, assume that previous synthesized frames are available. The condition imposed on the newly synthesized frame is that it is consistent with the instantaneous parallax computation and that it is smooth in time with respect to the previously generated frames. This

latter condition can be imposed by making a flow-based prediction based on the previous frames and making the difference from that prediction part of the error term. Similarly, if a sequence has already been generated, then the parallax-based frame (i.e., the warped high-resolution image) can be compared with the flow based temporally interpolated frame. This comparison can be used either to detect problem areas or to refine the parallax computation. This approach can be used without making rigidity assumptions or in conjunction with a structure/parallax constraint. In this latter case, the flow-based computation can operate with respect to the residual motion after the rigid part has been compensated. An extension of this is to apply the planar constraint across frames along with the global rigid motion constraint across all the pixels in one frame.

Additionally, a method of computing residual parallax 306 which avoids a potential problem with instability in the synthetic stereo sequence in three dimensional structure composed using the synthetic image 114 is to limit the amount of depth change between frames. To reduce this problem, it is important to avoid temporal fluctuations in the extracted parallax structure using temporal smoothing. A simple form of this smoothing can be obtained by simply limiting the amount of change introduced when updating a previous estimate. To do this in a systematic way requires inter-frame motion analysis as well as intra-frame parallax computation to be performed.

Occlusion detection 308 is helpful in situations in which an area of the view to be synthesized is not visible from the position of the high-resolution camera. In such situations, it is necessary to use a different source for the image information in that area. Before this can be done, it is necessary to detect that such a situation has occurred. This can be accomplished by comparing results obtained when image correspondence is computed bi-directionally. That is, in areas in which occlusion is not a problem, the estimated displacements from computing right-left correspondence and from computing left-right correspondence agree. In areas of occlusion, they generally do not agree. This leads to a method for detecting occluded regions. Occlusion conditions can also be predicted

from the structure of the parallax field itself. To the extent that this is stable over time areas of likely occlusion can be flagged in the previous frame. The bi-directional technique can then be used to confirm the condition.

Motion analysis 310 also improves the parallax computation of step 212. Motion analysis 310 involves analyzing frame-to-frame motion within the captured sequence. This information can be used to solve occlusion problems because regions not visible at one point in time may have been visible (or may become visible) at another point in time. Additionally, the problem of temporal instability can be reduced by requiring consistent three-dimensional structure across several frames of the sequence.

Analysis of frame-to-frame motion generally involves parsing the observed image change into components due to viewpoint change (i.e., camera motion), three dimensional structure and object motion. There is a collection of techniques for performing this decomposition and estimating the respective components. These techniques include direct camera motion estimation, motion parallax estimation, simultaneous motion and parallax estimation, and layer extraction for representation of moving objects or multiple depth surfaces. A key component of these techniques is the "plane plus parallax" representation. In this approach, parallax structure is represented as the induced motion of a plane (or other parametric surface) plus a residual per pixel parallax map representing the variation of induced motion due to local surface structure. Computationally, the parallax estimation techniques referred to above are essentially special cases of motion analysis techniques for the case in which camera motion is assumed to be given by the fixed stereo baseline.

To improve processing efficiency, the parallax computation (or flow computation) can be performed at the resolution of the low resolution image. Then, the parallax information can be projected to generate a correspondence map at the higher resolution. The subsequent image warping and/or compositing process is then performed using the projected parallax information.

Once the parallax field has been computed in step 212, it is used to produce the high-resolution synthesized image 114 in step 214. The compositing and warping step 214 is depicted in Fig. 2 and in greater detail in Fig. 4. Conceptually this process involves two steps: parallax interpolation and image warping. In practice these two steps are usually combined into one operation as represented by step 214. In either case, for each pixel in the to-be-synthesized image, the computation of step 214 involves accessing a displacement vector specifying a location in the high-resolution source image from the high resolution camera 206 (step 502), accessing the pixels in some neighborhood of the specified location and computing, based on those pixels (step 504), an interpolated value for the synthesized pixels that comprise the synthetic image 114 (step 506). Step 214 should be performed at the full target image resolution. Also, to preserve the desired image quality in the synthesized image 114, the interpolation step 506 should be done using at least a bilinear or bicubic interpolation function.

Even more effective warping and compositing algorithms can make use of motion, parallax, other information (step 508). For example, the location of depth discontinuities from the depth recovery process can be used to prevent spatial interpolation in the warping across such discontinuities. Such interpolation can cause blurring in such regions. In addition, occluded areas can be filled in with information from previous or following frames using flow based warping. The technique describe above in the discussion of plane plus parallax is applicable for accomplishing step 508.

Also, temporal scintillation of the synthesized imagery can be reduced using flow information to impose temporal smoothness (step 510). This flow information can be both between frames in the synthesize sequence, as well as between the original and synthesized imagery. Scintillation can also be reduced by adaptively peaking pyramid-based appearance descriptors for synthesized regions with the corresponding regions of the original high resolution frames. These can be smoothed over time to reduce "texture flicker."

The compositing and warping step 214 can also be performed using data collected over an image patch, rather than just a small neighborhood of pixels. For example, the image can be split up into a number of separate regions, and the resampling is performed based on the area covered by the region in the target image (step 512).

The depth recovery may not produce completely precise depth estimates at each image pixel. This can result in a difference between the desired intensity or chroma value and the values produced from the original high-resolution imagery. The warping module can then choose to select one or more of the following options as a correction technique (step 514), either separately, or in combination:

- leave the artifact as it is (step 516)
- insert data that has been upsampled from the low-resolution imagery (step 518)
- use data that has been previously synthesized (step 520)
- allow an operator to manually correct the problem (step 522).

A Just Noticeable Difference (JND) technique can be used for selecting the appropriate combination of choices. The JND measure is performed on the synthesized sequence by comparing the difference between a low-resolution form of the synthesized data and data from the low-resolution camera to create a JND map representing a quality-of-parallax computation measure. Various JND measures are described in United States Patent Application No.'s 09/055,076, filed April 3, 1989, 08/829,540, filed March 28, 1997, 08/829,516, filed March 28, 1997, and 08/828,161, filed March 28, 1997 and United States Patent No.'s 5,738,430 and 5,694,491, all of which are incorporated herein by reference in their entireties. Additionally, the JND can be performed between the synthesized high-resolution image data, and the previous synthesized high-resolution image after being warped by the flow field computed from the parallax computation in step 212.

Once the high-resolution synthetic image is created for the low resolution viewpoint, the original high resolution image and the synthetic image can be used to form a high resolution stereo image.

Although the embodiment which incorporate the teachings of the present invention have been shown and described in detail herein, those skilled in the art can readily devise many other varied embodiments that still incorporate these teachings and spirit of the invention.

What is claimed is:

1. Apparatus for generating a synthetic image comprising:

a first camera producing at least one image;

at least one additional camera that differs from the first camera by at least one characteristic or parameter; and

an image processor coupled to said first and at least one additional camera, for transforming said at least one image from the first camera to other imagery where one or more of the characteristics or parameters are different.

2. The apparatus of claim 1 where the said other imagery is transformed by at least one characteristic or parameter of one or more of the additional cameras.

3. The apparatus of claim 1 where the characteristic and parameter comprises spatial position of the camera.

4. The apparatus of claim 1 where the characteristic or parameter comprises spatial resolution of the camera.

5. The apparatus of claim 1 where the characteristic or parameter comprises spectral characteristics of the camera.

6. The apparatus of claim 1 where the characteristic or parameter comprises spatial layout of the coordinate system of the camera.

7. A method for synthesizing an image comprising the steps of:

supplying first resolution images recorded from a first resolution camera;

supplying second resolution images recorded from a second resolution camera, where said first resolution is greater than said second resolution;

computing image flow using a plurality of images; and

warping said first resolution image to the viewpoint of said second resolution camera using said image flow t produce a synthetic image.

8. The method of claim 7 further comprising the step of:

computing a parallax estimation for said low-resolution image; and

projecting the parallax estimation for use in computing the synthetic image.

9. The method of claim 7 wherein said step of computing said parallax estimation further comprises the step of:

enhancing said parallax computation by performing one or more augmentation routines selected from the group consisting of:

dividing the images into tiles, correlating depth, performing Just Noticeable Differences techniques, checking correspondences, and applying techniques to avoid blank areas.

10. A computer-readable medium having stored thereon a plurality of instructions, the plurality of instructions including instructions which, when executed by a processor, cause the processor to perform the steps comprising of:

supplying first resolution images recorded from a first resolution camera;

supplying second resolution images recorded from a second resolution camera, where said first resolution is greater than said second resolution;

computing image flow using a plurality of images; and

warping said first resolution image to the viewpoint of said second resolution camera using said image flow t produce a synthetic image.

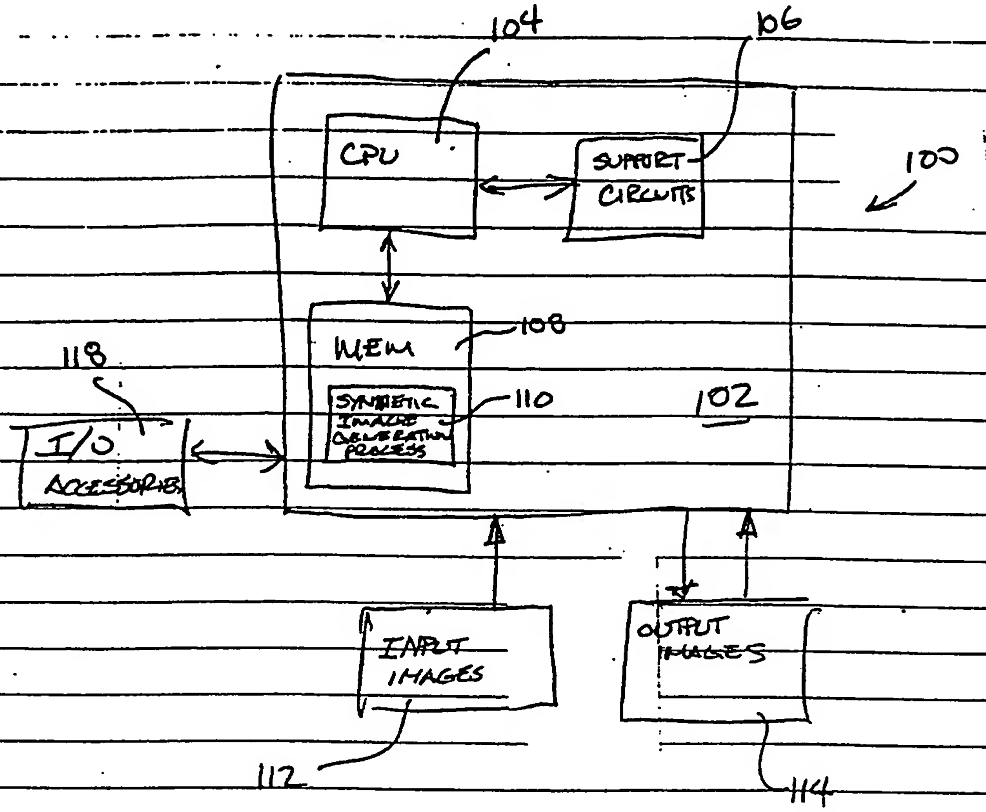

104
106
CPU
SUPPORT CIRCUITS
100
MEM
108
118
SYNTHETIC IMAGE GENERATION PROCESS
110
I/O ACCESSORIES
102
INPUT IMAGES
OUTPUT IMAGES
112
114

FIG 1

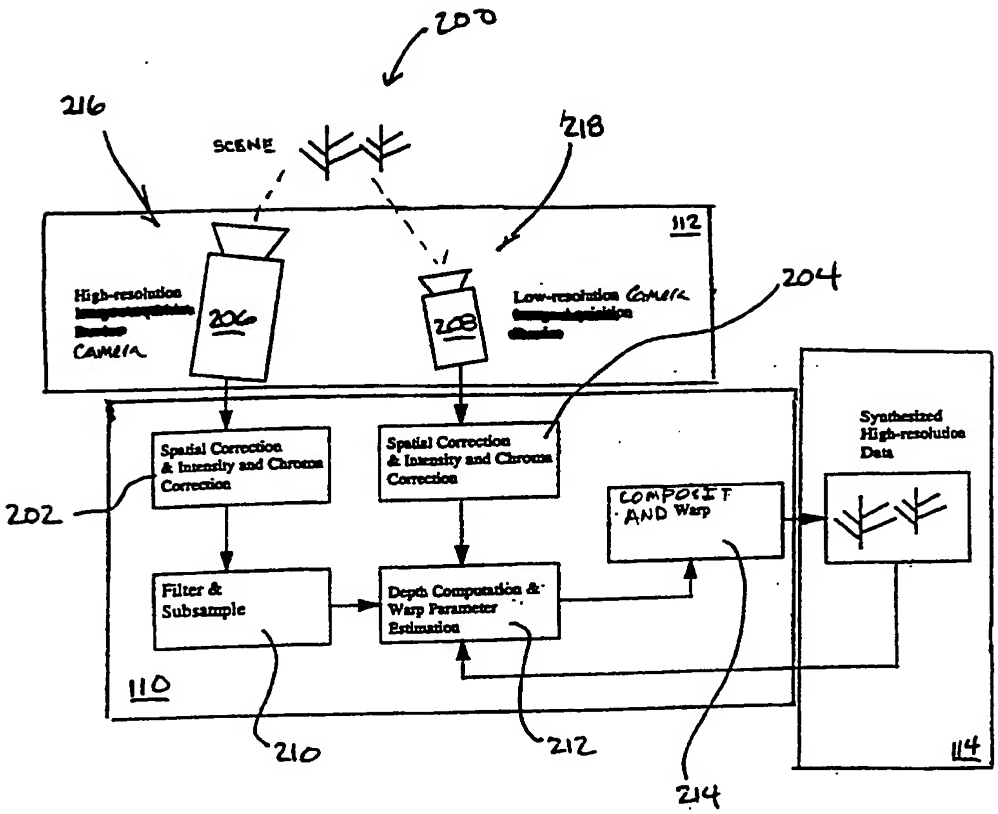

200
216
SCENE
218
112
High-resolution [redacted] CAMERA
206
208
Low-resolution CAMERA [redacted]
204
Spatial Correction & Intensity and Chroma Correction
Spatial Correction & Intensity and Chroma Correction
202
COMPOSIT AND Warp
Synthesized High-resolution Data
Filter & Subsample
Depth Computation & Warp Parameter Estimation
110
210
212
214
114

FIG 2

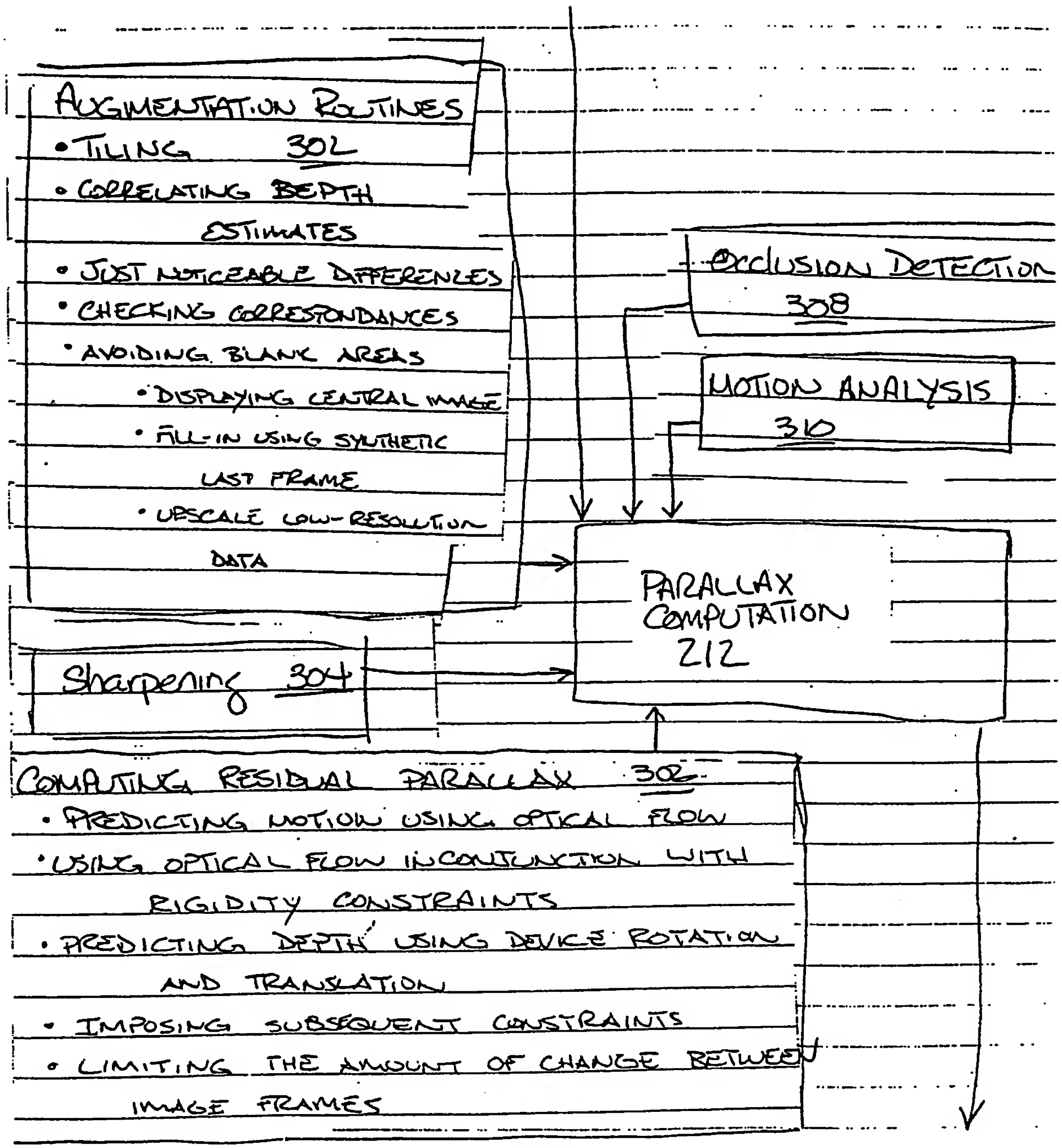

AUGMENTATION ROUTINES
• TILING 302
• CORRELATING DEPTH ESTIMATES
• JUST NOTICEABLE DIFFERENCES
• CHECKING CORRESPONDANCES
• AVOIDING BLANK AREAS
• DISPLAYING CENTRAL IMAGE
• FILL-IN USING SYNTHETIC LAST FRAME
• UPSCALE LOW-RESOLUTION DATA
Occlusion Detection 308
MOTION ANALYSIS 310
PARALLAX COMPUTATION 212
Sharpening 304
COMPUTING RESIDUAL PARALLAX 306
• PREDICTING MOTION USING OPTICAL FLOW
• USING OPTICAL FLOW IN CONJUNCTION WITH RIGIDITY CONSTRAINTS
• PREDICTING DEPTH USING DEVICE ROTATION AND TRANSLATION
• IMPOSING SUBSEQUENT CONSTRAINTS
• LIMITING THE AMOUNT OF CHANGE BETWEEN IMAGE FRAMES

FIG 3

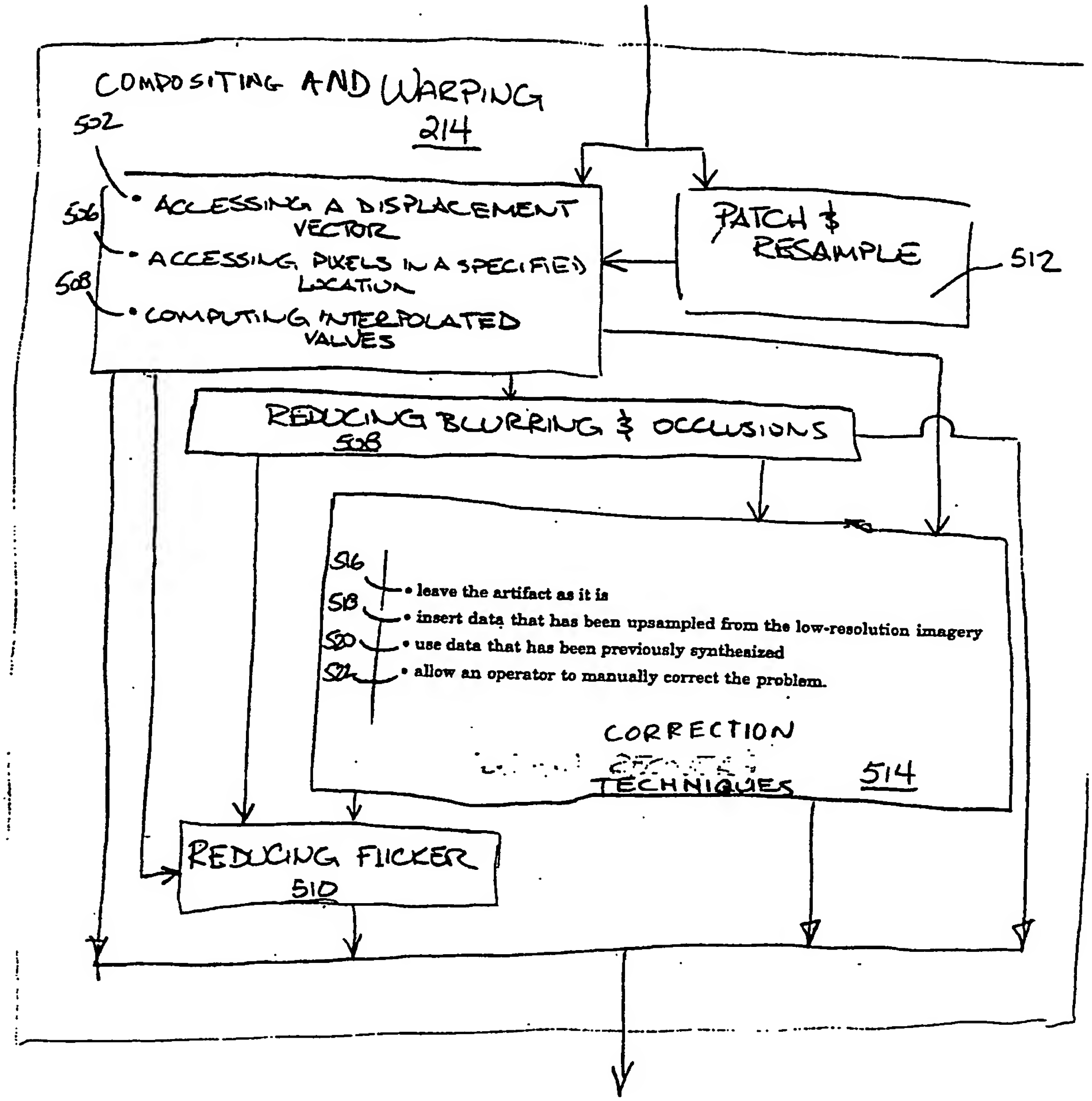

COMPOSITING AND WARPING
214
502
506
508
• ACCESSING A DISPLACEMENT VECTOR
• ACCESSING PIXELS IN A SPECIFIED LOCATION
• COMPUTING INTERPOLATED VALUES
PATCH & RESAMPLE
512
REDUCING BLURRING & OCCLUSIONS
508
516
518
520
522
• leave the artifact as it is
• insert data that has been upsampled from the low-resolution imagery
• use data that has been previously synthesized
• allow an operator to manually correct the problem.
CORRECTION
TECHNIQUES
514
REDUCING FLICKER
510

FIG 4

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/19706

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04N13/02 G06T3/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04N G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 21690 A (SARNOFF CORP) 22 May 1998 (1998-05-22)	1-3,5,6
A	page 3, line 4 -page 4, line 24; figures 1,2	7,10
A	EP 0 641 132 A (MATSUSHITA ELECTRIC IND CO LTD) 1 March 1995 (1995-03-01)	

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 November 1999	29/11/1999
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer De Paepe, W

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 99/19706

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
WO 9821690	A	22-05-1998	NONE			
EP 0641132	A	01-03-1995	DE	69417824	D	20-05-1999
			DE	69417824	T	12-08-1999
			EP	0888017	A	30-12-1998
			JP	7167633	A	04-07-1995
			US	5726704	A	10-03-1998
			US	5801760	A	01-09-1998